Supporting Information

Quick guide to GoDig

Qing Yu,[1] Xinyue Liu,[1] Mark P Keller,[2] Jose Navarrete-Perea,[1] Tian Zhang,[1] Sipei Fu,[1] Laura P Vaites,[1] Steven R Shuken,[1] Ernst Schmid,[3] Gregory R Keele,[4] Jiaming Li,[1] Edward L Huttlin,[1] Gary A Churchill,[4] Devin K Schweppe,[5] Alan D Attie,[2] Joao A Paulo,[1] Steven P Gygi[1,*]

[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA.
[2]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA
[3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.
[4]The Jackson Laboratory, Bar Harbor, Maine 04609, USA
[5]Department of Genome Sciences, University of Washington, Seattle, WA 98105, USA

Download Tomahto

- Prerequisite: Thermo API license (https://github.com/thermofisherlsms/iapi)
- Download website: https://gygi.med.harvard.edu/software (also smarttmt.org).
- Orbitrap Tune software: 3.5 or later version

Smart-TMT

Real-Time-Search

The Real-Time Search** (RTS-MS^3) provides real-time (<5 ms / spectrum) spectral identification and triggers SPS-MS^3 scans that utilize assigned and pure fragment ions for accurate quantitation.Time consuming SPS-MS^3 spectra are only acquired after confident peptide identification, greatly increasing the number of peptides interrogated and reducing the effects of isobaric interference. **Try it!**

Tomahto

The Tomahto software provides real-time instrument control and decision making. Tomahto enables simplied implementation of TOMAHAQ targeted assay. It provides an array of functionalities including MS1 peak detection, MS2 real-time peak matching (RTPM), MS2 fragmentation pattern match, SPS ion purity filter, MS3 automatic gain control (AGC), MS3 quant scan insertion, and target peptide close-out. In addition to controlling data acquisition, it also allows real-time data visualization and post-acquisition analysis. **Try it!**

GoDig

The GoDig software is a real-time analytics platform that enables next-generation TMT-based multiplexed targeted proteomics. It obviates the need for making internal peptide standards and tedious method curation. As a result, it essentially makes targeting nearly any prviously detected peptides possible. One only needs a list of peptides and GoDig will perform real-time elution calibration and spectral matching to identify the targets, and prompt SPS-MS3 scans to quantify them down to attomole level. **Try it!**

General considerations for trigger peptide selection

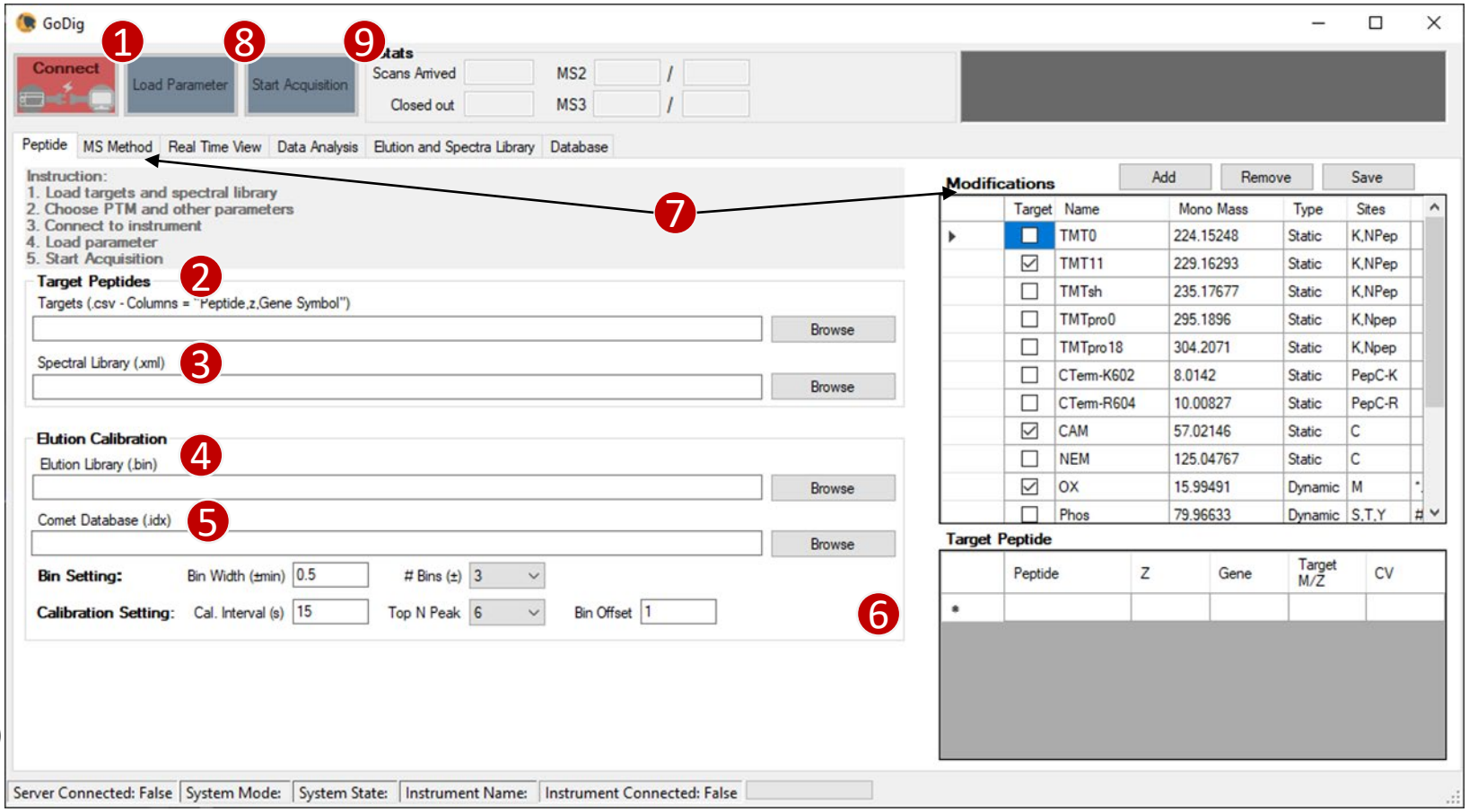
1. Should be unique to the protein of interest
2. Should avoid missed cleavages to minimize artifact caused by sample preparation
3. Should not contain methionine to avoid artifact caused by oxidation
4. If a peptide contains cysteines, an extra step needs to be taken to reduce/alkylate the free cysteines using the same protocol as the endogenous protein sample preparation
5. Preferred peptide length is 8~20 amino acids to ensure proper hydrophobicity
6. We prefer to have 2 peptides per target protein as a fail-safe measure

How to acquire data using GoDig.exe

1. Connect instrument
2. Select target peptide list (.csv)

Example of input .csv file

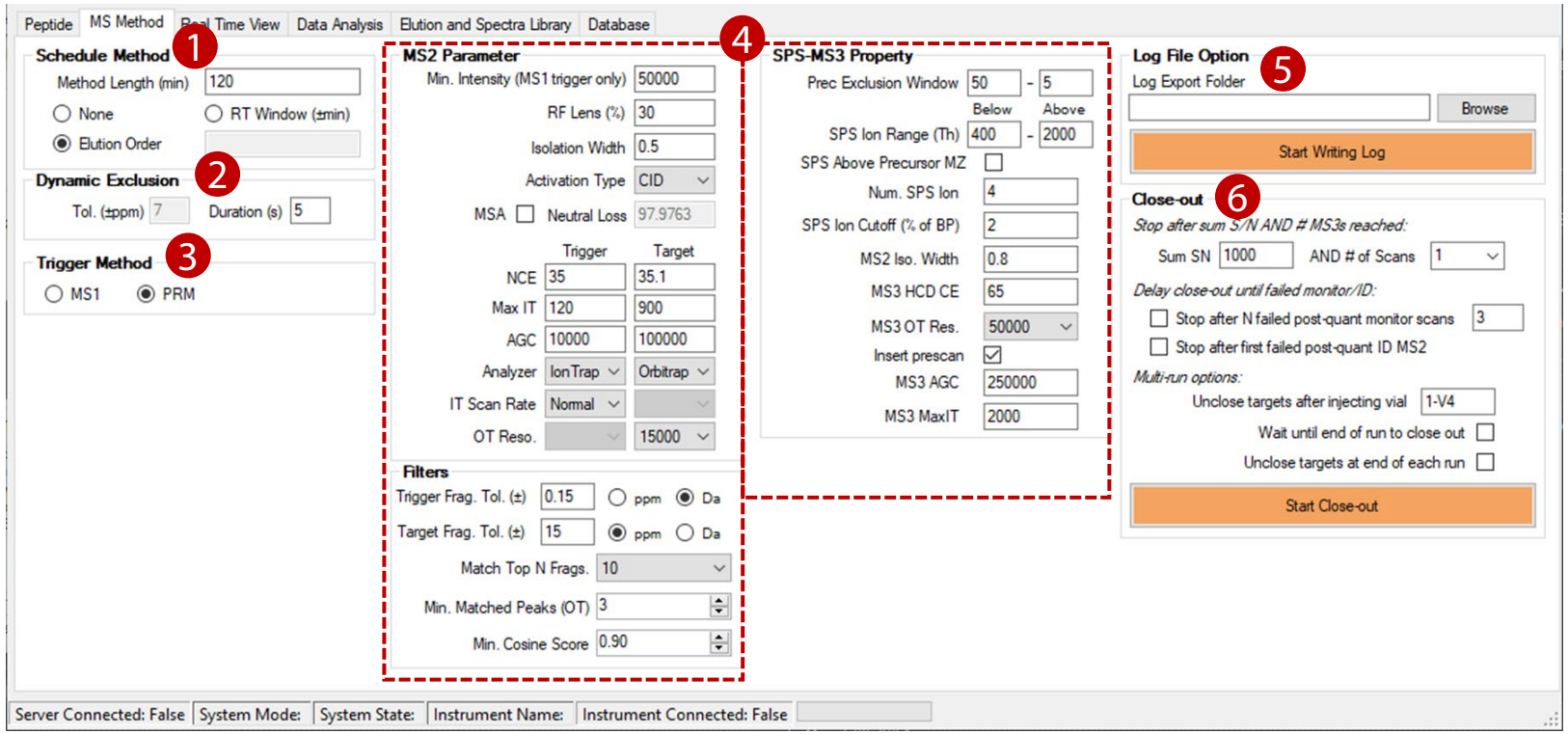
GeneSymbol	Peptide
Abhd1	YTSVAFGYK
Abhd1	VLDVDFAIK
Abhd10	SELPNLAYK
Abhd10	IPYSFIK
Abhd11	LNLDTLAQHLDK
Abhd11	TNFNSLAK
Abhd11	VLTVDAR
Abhd12	VPYFIDLK



3. Select peptide spectral library (.xml)
4. Select peptide elution library
   (e.g., demo_elution_library_4_cell_line_tmtpro.bin)
5. Select proteome database
   (.idx; e.g., demo_uniprot_HUMAN_TrypsinKR_TMTPro.fasta.idx; see next slide about how to build one)
6. Select elution calibration interval and desired bin width
7. Select peptide modification and data acquisition parameters (see next slide)
8. Load parameter. This may take a couple minutes.
9. Start acquisition

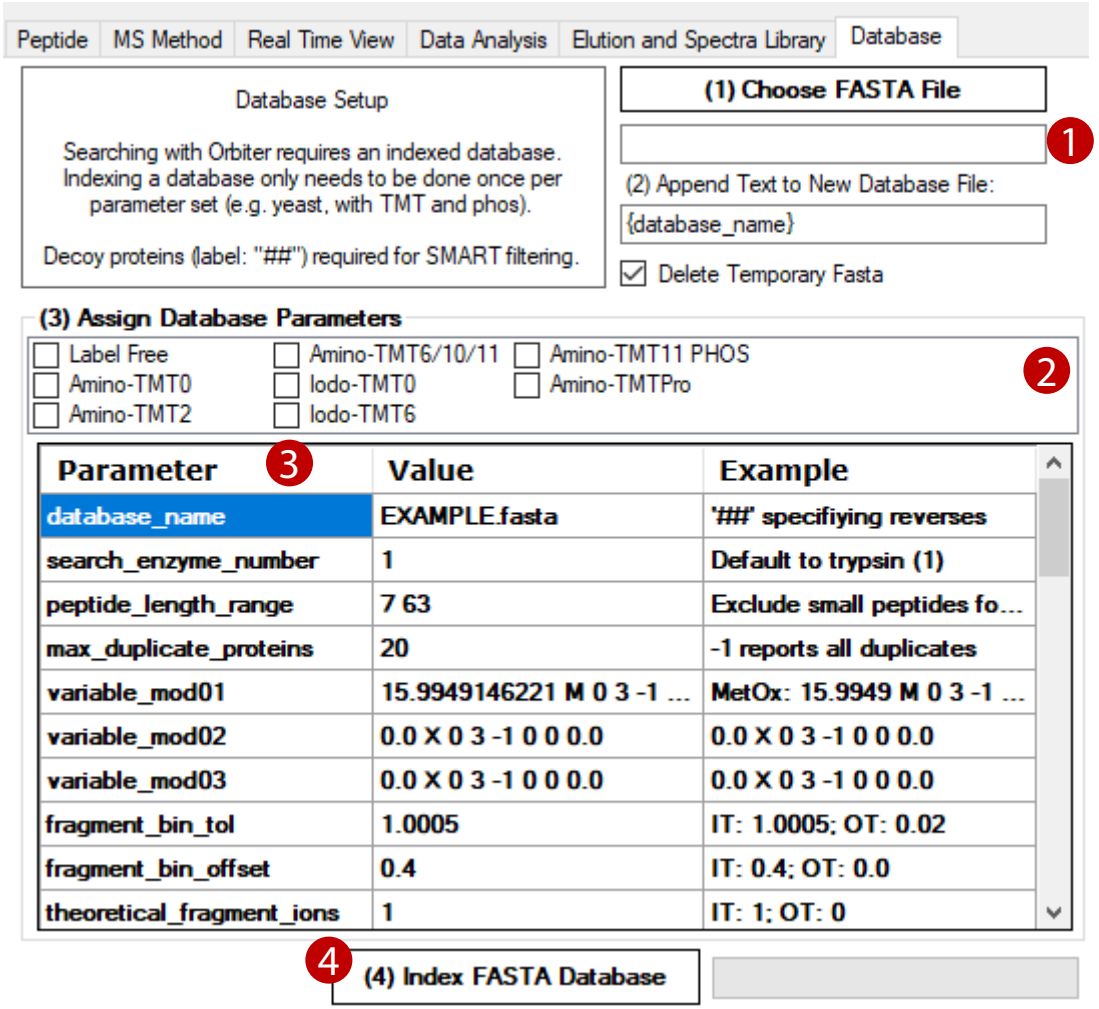
How to acquire data using GoDig.exe

1. Select scheduling method (default to elution order)
2. Set dynamic exclusion (how frequent a target should be monitored)
3. Set trigger method (default to PRM)
4. Set MS parameters
5. Choose directory to write log file
6. Choose close-out option if needed



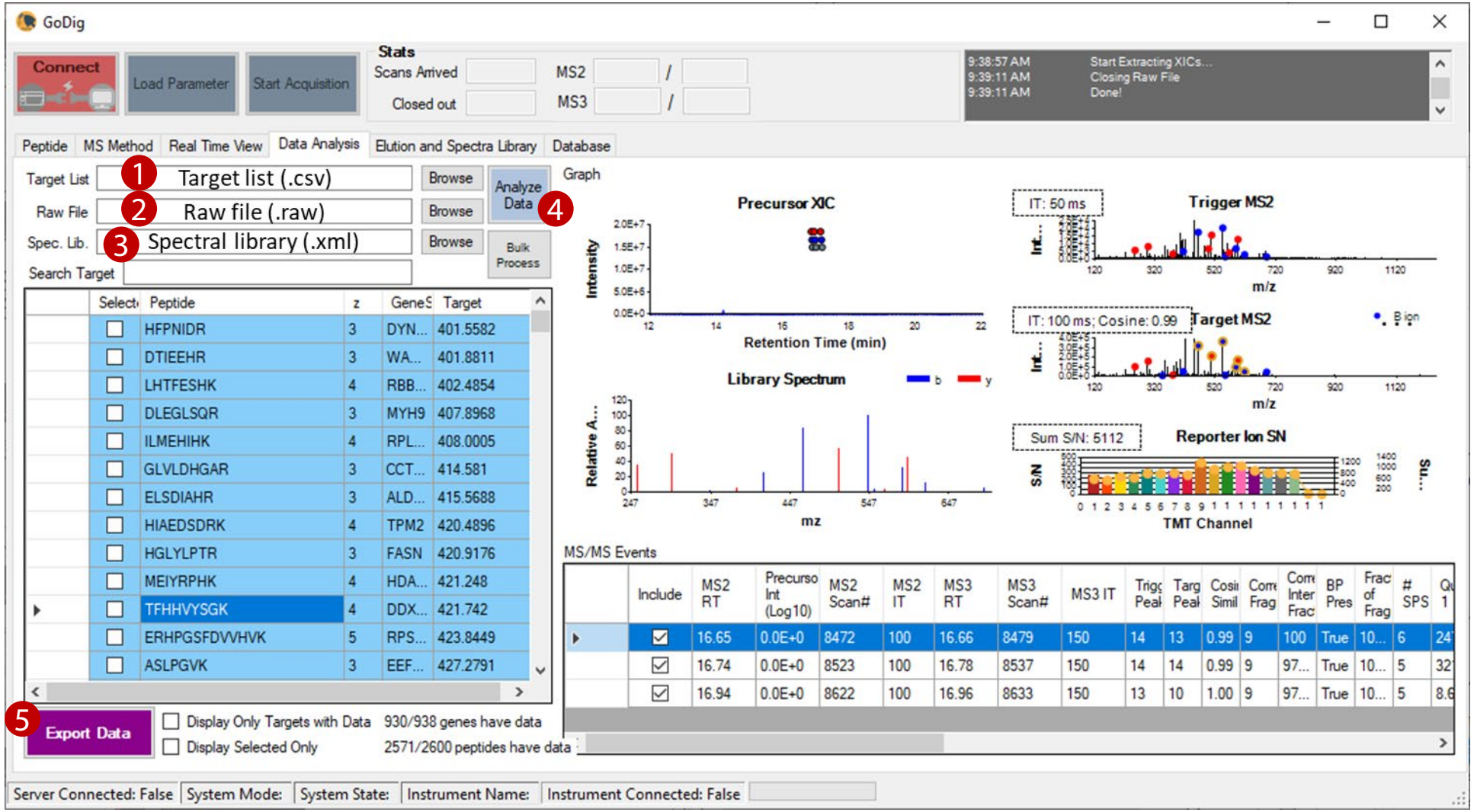
How to index a proteome .fasta file to be used for elution calibration

1. Select proteome fasta file
   (can be obtained from Uniprot https://www.uniprot.org/)
2. Select TMT labeling
3. Select search parameters. GoDig uses Comet as its real-time search engine. Please refer to Comet website for all details regarding the parameters (https://comet-ms.sourceforge.net/parameters/)
4. Index database. A .idx file will be generated in the same folder as the .fasta



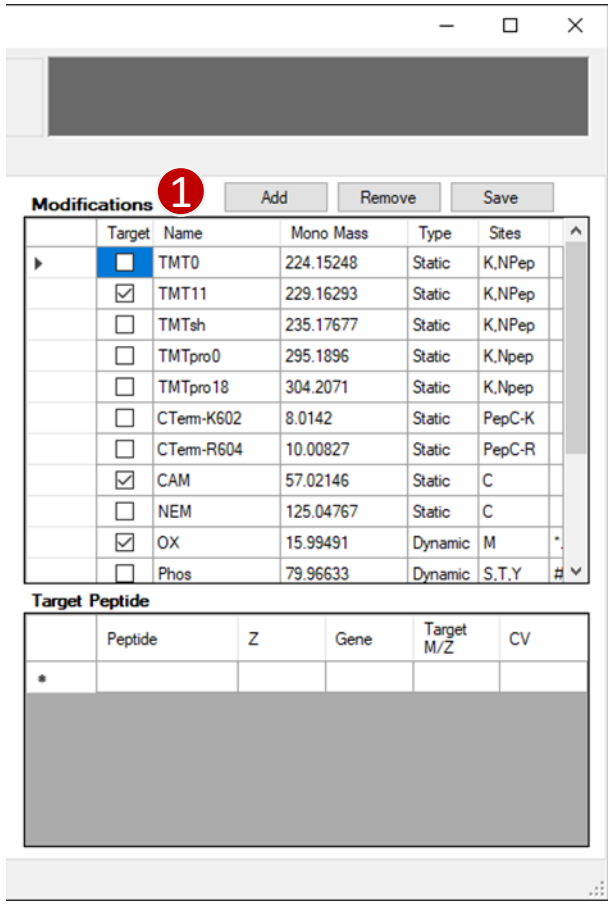
Data analysis

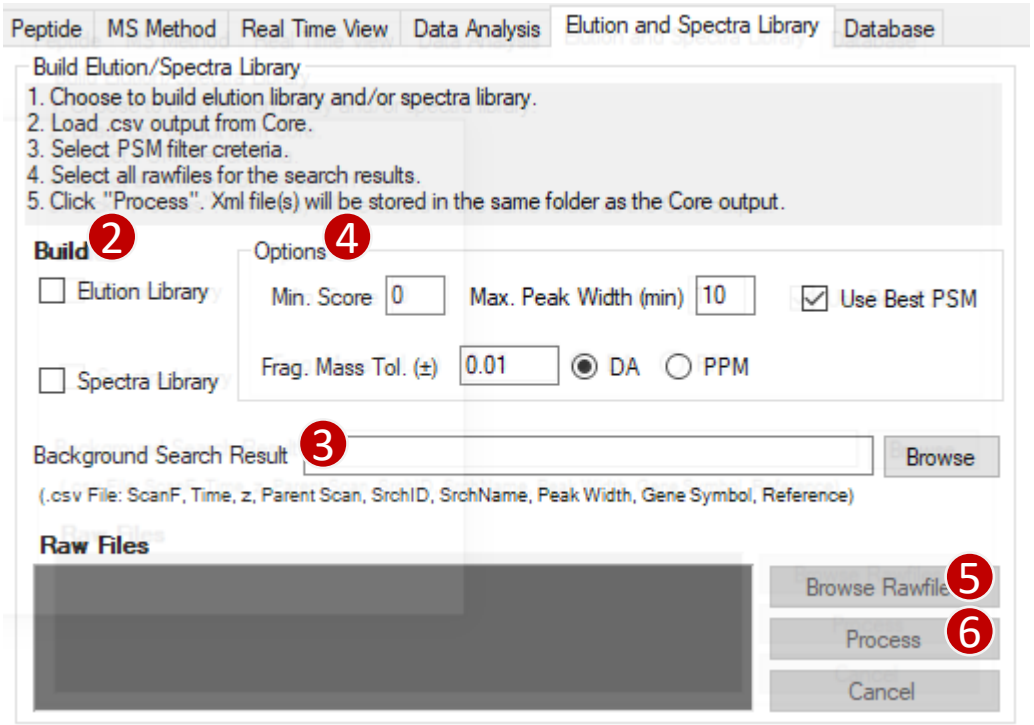
1. Select peptide list (.csv) and raw file.
   Note: you have to also select same modification, and fragment match tolerance in the method editor as the rawfile was acquired with.
2. Select raw file (.raw)
3. Select spectra library (.xml)
4. Process
5. Export to .csv



How to build custom library

1. Select peptide modification
2. Select which library to build (e.g., spectral and elution)
3. Select the .csv file that contains all identification information to be used for the library (see next slide for details)
4. Additional filters to remove low score peptides or peptide with large elution window. Fragment mass tol need to be set to build spectral library.
5. Raw files that are used to generate the .csv search result
6. Start to build (this may take a while, depending on the size of the dataset)





How to build custom library--.csv file

Example csv input:

Peptide	z	Theo m/z	Parent Scan	Reference	ScanF	Expect	Time	SrchID	SrchName	Peak Width	Gene Symbol	PepID
R.QQEGESR.L	2	569.2946	786	sp|P00338-3|LDHA_HUMAN	795	6.592	5.3	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3003	LDHA	43
R.SQEAGGR.V	2	504.7733	786	sp|P48509|CD151_HUMAN	800	5.252	5.31	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3003	CD151	48
K.AGQGSGSR.K	2	512.2787	794	sp|A4FU01|MTMRB_HUMAN	808	6.686	5.33	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3093	MTMR11	55
K.NQDDADR.A	2	569.2764	794	sp|O75937|DNJC8_HUMAN	813	7.759	5.33	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.287	DNAJC8	60
K.REQAEEER.Y	3	450.9026	817	sp|Q9UII2|ATIF1_HUMAN	826	5.471	5.36	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.2783	ATP5IF1	71
K.RAQEEAER.L	3	431.5674	817	tr|E7EQR4|E7EQR4_HUMAN	827	6.342	5.36	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3003	EZR	72
K.NEEDEGHSNSSPR.H	3	587.9368	817	sp|Q14103|HNRPD_HUMAN	830	11.026	5.36	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.2565	HNRNPD	75
R.CNDQDTR.T	2	606.7835	817	sp|P02751-15|FINC_HUMAN	833	8.64	5.37	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.2565	FN1	78
K.NASCGTR.S	2	535.2726	825	sp|P60468|SC61B_HUMAN	838	4.764	5.37	540183	ec04297_qy_GoDig_4cell_CIDhrMS2	0.3093	SEC61B	82

Peptide: peptide sequence with or without two flanking residues (e.g., R.QQEGESR.L or QQEGESR)
z: peptide charge state
Theo m/z: theoretical peptide m/z
Parent Scan: master scan the PSM is based on
Reference: Uniprot reference
ScanF: Scan number of the PSM
Expect: Expect score from Comet search or XCorr from Sequest search (Can be other score too but need to use the same column header)
Time: Scan time
SrchName: raw file name without the file suffix (e.g., ec04297_qy_GoDig_4cell_CIDhrMS2 from ec04297_qy_GoDig_4cell_CIDhrMS2.raw)
Peak Width: peptide peak width
Gene Symbol: gene symbol
PepID: each PSM from one raw file has a unique peptide ID. If not provided by a search engine, **ScanF** can be used instead since it is unique in each raw file.

More questions?

Please email Qing Yu (qing_yu2@hms.harvard.edu) or Steven Gygi (steven_gygi@hms.harvard.edu)